10-2000

Two Large-Insert Soybean Genomic Libraries Constructed in a Binary Vector: Applications in Chromosome Walking and Genome Wide Physical Mapping

K. Meksem
Southern Illinois University, meksemk@siu.edu

K. Zobrist
Southern Illinois University

E. Ruben
Southern Illinois University

D. L. Hyten
Southern Illinois University, david.hyten@unl.edu

T. Quanzhou
Texas A&M University

See next page for additional authors

Published in *Theoretical and Applied Genetics* 101:5–6 (October 2000), pp. 747–755;
doi: 10.1007/s001220051540

Submitted May 7, 1999; accepted February 18, 2000.

Two Large-Insert Soybean Genomic Libraries Constructed in a Binary Vector: Applications in Chromosome Walking and Genome Wide Physical Mapping

K. Meksem,[1] K. Zobrist,[1] E. Ruben,[1] D. Hyten,[1] T. Quanzhou,[2] H-B. Zhang,[2] and D. A. Lightfoot[1]

1. Department of Plant Soil and General Agriculture, Southern Illinois University at Carbondale, Room 176, Carbondale, IL 62901–4415, USA
2. Department of Soil and Crop Sciences, Crop Biotechnology Center, Texas A & M University, College Station, TX 77843, USA

Corresponding author – K. Meksem, email http://www.coalab.siu.edu/genome/meksemk@siu.edu

Abstract
Large DNA insert libraries in binary T-DNA vectors can assist in the isolation of the gene(s) underlying a quantitative trait locus (QTL). Binary vectors facilitate the transfer of large-insert DNA fragments containing a QTL from *E. coli* to *Agrobacterium* sp. and then to plants. We constructed two soybean large-insert libraries from cv. Forrest in the pCLD04541 (V41) binary vector after partial digestion of genomic high-molecular-weight DNA with *Bam*HI or *Hin*dIII. The libraries contain 76,800 clones with an average insert size of 125 kb, and therefore represent 9.5-fold haploid genome equivalents. Colony hybridization using a chloroplast-specific probe infers that the libraries contain less than 0.5% clones of chloroplast DNA origin. These two libraries have provided clones for physical mapping of the soybean genome and for isolation of a number of disease resistance genes. One microsatellite marker was identified from the clone that hybridized to the Bng122 RFLP probe. The sequence-tagged site was used for genetic mapping and marker-assisted selection for genes underlying resistance to the soybean cyst nematode and sudden death syndrome.

Keywords: bacterial artificial chromosome, soybean, physical mapping, targeted microsatellites, positional cloning, disease resistance, genomics, cyst nematode

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the world's major crops. Soybean genomics is complicated by the complexity of the genome and the abundance of homeologous regions derived from genome duplication (Shoemaker et al. 1996). In the past decade, the development of DNA markers has facilitated the construction of genetic linkage maps of soybean. Maps based on restriction fragment length polymorphism (RFLP) (Lark et al. 1993; Shoemaker and Specht. 1995), random amplified polymorphic DNA (RAPD), (Chang et al. 1997), ámplified fragment length polymorphism (AFLP, Keim et al. 1997) and microsatellite (Cregan et al. 1999) markers have been developed and integrated. These maps have allowed the identification of many economically important genes conditioning qualitative traits (Keim et al. 1990; Diers et al. 1992) and quantitative trait loci (QTLs) (Mansur et al. 1993; Webb et al.1995; Chang et al. 1996; Lee et al. 1996; Concibidio et al. 1997; Mian et al. 1997; Tamulonis et al. 1997). Fine mapping of a few economically important loci has been achieved (Kolchinsky et al. 1997; Ashfield et al. 1998; Matthews al. 1998; Meksem et al. 1999).

Positional cloning is a promising method for isolating and studying genes for which only the locus-derived phenotype is known (Martin et al. 1993; Bent et al. 1994; Mindrinos et al. 1994; Grant et al. 1995; Song et al. 1995). Large-insert libraries constructed in YACs (yeast artificial chromosome) and BACs (bacterial artificial chromosome) have been the tools of choice for generating high-resolution physical maps and identifying candidate genes during positional cloning (Arondel et al. 1992; Tanksley et al. 1995). BAC libraries are simpler to construct, manipulate and maintain than YAC libraries (Frijters et al. 1997; Yang et al. 1997) and do not suffer problems associated with homologous recombination and differential colony growth. BAC libraries have been constructed for several plant species (Woo et al. 1994; Wang et al. 1995; Frijters et al. 1996; Zhang et al. 1997; Bent et al. 1998; Vinatzer et al. 1998) including soybean (Marek and Shoemaker 1997, Danesh et al. 1998, Salimath and Bhattacharyya 1999). During positional cloning, large-insert libraries are often used to integrate fine genetic maps with physical maps. However, the generation of new genetic markers for chromosome walking is a challenge in complex genomes due to the repeat DNA, homeologous sequences and the rarity of polymorphism (Danesh et al. 1998).

Genes underlying QTL, or with related functions such as disease resistance, seem to be organized in clusters (Staskawicz et al.1995). BAC vectors are designed for cloning large fragments that might contain a gene cluster or an intact locus. Therefore, the ability to introduce large insert DNA clones into plant cells (Hamilton et al. 1996) provided potential for functional genomics, genetic engineering of complex loci and the assembly of several unlinked genes into a single locus. However, there are only a few reports of transformation of large DNA fragments (>70 kb) into a plant genome (Hamilton et al. 1996; Liu et al. 1999). Although soybean cells can be successfully transformed using a number of different methods (Rong et al. 1996), it has been recalcitrant to transformation due to a low efficiency in

the regeneration of transformed cells into fertile plants (Singh et al. 1998). Therefore, functional genomics and xenogenetics can provide more targets for soybean transformation than can be analyzed without large-insert transformation methods.

The soybean cultivar Forrest represents a source of genes for resistance to damaging diseases including cyst nematode (SCN), sudden death syndrome (SDS) (Chang et al. 1997) and for beneficial agronomic traits (Kilo et al. 1997; Njiti et al. 1997). We have constructed a high-resolution genetic map of three genomic regions that contain genes for resistance to root pathogens; *Rhg1*, and *Rhg4* that confers resistance to *Heterodera glycines* I. race 3 (SCN) and *Rfs1* that confers resistance *to Fusarium solani f.* sp. *glycines* (SDS) (Meksem et al. 1999).

This report describes the construction and characterization of two Forrest large-insert libraries with a combined 9.5-fold haploid genome equivalence and an average insert size of 125 kb in the pCLD04541 (V41) binary T-DNA vector. We developed methods to isolate microsatellite DNA markers from large-insert DNA clones. Targeted microsatellites are suitable for gap closure between large-insert DNA contigs, for positional cloning, and for marker-assisted selection.

Materials and methods

Plant materials

The soybean cultivar Forrest (Hartwig and Epps 1973) was used. Seeds were obtained from the SIUC (Southern Illinois University at Carbondale) field program. For DNA isolation, plants were grown in the greenhouse in soil for 14 days, with 16 h light, 8 h dark at 22°C. Two weeks after germination, seedlings were transferred into continuous dark for 3 days to reduce the carbohydrate content of the soybean leaves. This step is important as it lowers starch reserves.

Vectors

The vector V41 (pCLD04541) was developed by Dr. C. Dean (John Innes Institute, Norwich, UK; Jones et al. 1992) and kindly provided by Dr. H. Zhang, (Texas A & M University, USA). The V41 is a binary vector for *Agrobacterium*-mediated plant transformation (Bent et al. 1994) and has been shown to be capable of stable maintenance of large plant DNA fragments in *E. coli* (Quanzhou and Zhang 1998). The pGEM-®3Z vector was purchased from Promega (Madison, Wis., USA).

DNA probes and primers

The Bng122 RFLP probe was provided kindly by Dr. E. Vallejos (University of Florida, USA). The microsatellite primers (BARCSATT 309, BARC-SATT 275, BARC-SATT 163) for screening the BAC library were kindly provided by Dr. P. Cregan (USDA, Beltsville, Md., USA). The microsatellite primers for SIUCSAT122 were generated at SIUC from a BAC clone (Meksem et al. 1999). The pBLT65 primers were provided by Dr. B. Matthews (USDA, Beltsville Md., USA). The pA381 primers were provided by Dr. P. Gresshoff (University of Tennessee, Knoxville Tenn., USA), The NBS5 was provided by Dr. M.A. Saghai Maroof (Virginia Polytechnic Institute and State University, Blacksburg, Va., USA) (Yu et al. 1996).

The (AT)10, (ATT)10, (AAT)10, (CT)10, (AG)10, (CT)10, (AC)10, (CG)10, (CC)10, (GG)10 oligomer repeats were purchased from Research Genetics (Huntsville, Ala., USA). The sequences of other primers used are shown in table 1.

Table 1 Sequences of the primers used for the large insert DNA library screening and genetic mapping

Marker	Forward: 5′>3′	Reverse: 5′>3′
BARC-Satt309	GCGCCTTCAAATTGGCGTCTT	CGCCTTAAATAAAACCCGA-AACT
BARC-Satt275	GCGGGATAATTGGTTTTACGA-AAATGC	GCGCCTAATCACCTAAAAA-AACGTTTA
BARC-Satt163	AATAGCACGAGAAAAGGAGA-GA	GTGTATGTGAAGGGGAAAA-ACTA
SIUC-Sat122	CTCACAAAATTGAAATGTATC	CCTTTTTCATCTTGAAAAT
pBLT65	CAACAGTTGGGACTAAGACT	CTAACACCAGGAGTGCTTGC

Preparation and digest of high-molecular-weight DNA

The method for high-molecular-weight (HMW) DNA preparation from plant nuclei is described by Zhang et al. (1995). Nuclei were prepared from 25 g of leaves and embedded in 12 ml of 0.5% (w/v) low-melting-point agarose microbeads or plugs. Partial digestion of the microbeads (100 μl contained about 1 μg DNA) was performed by fixing the time of digestion to 10 min at 37°C and varying the concentration of the enzyme *Bam*HI (Gibco-BRL, Md.) over five steps from 0.1 to 2 U per 1μg of DNA. When the plugs were used (one plug contained about 1μg of DNA), each plug (volume about 70 μl) was cut into 12 pieces, the digestion time was extended to 20 min at 37°C and the amount of the enzyme *Hin*dIII (Gibco-BRL, Md.) ranged from 0.1–2 units per 1 μg of DNA (as modified from Bonnema et al. 1996). Each digestion was carried out in a total volume of 200 μl. Reactions were stopped by adding 20 μl of ice-cold 0.5 *M* EDTA pH 8, on ice. DNA in microbeads is accessible to the restriction enzyme and easy to pipette due to the small bead size but does not yield the same amounts of DNA between experiments. The *Hin*dIII library was made from plugs that have a constant yield of DNA but reduce enzyme access. The plugs were cut to small pieces to improve access of the restriction enzymes to the DNA.

Construction, storage and analysis of the large-insert libraries

The vector V41 was purified with Qiagen plasmid purification kit (Qiagen, Hilden, Germany), followed by two cesium chloride gradients. The vector was digested (with *Bam*HI or *Hin*dIII) and then dephosphorylated to prevent self ligation. Partially digested HMW DNA was size-selected on 1% (w/v) pulsed field low-melting-point agarose gels in 0.5 × TBE (45 m*M* Trizma base, 45 mM boric acid, 1 mM EDTA, pH 8.3) by PFGE (pulsed field gel electrophoresis) on a CHEF DRIII (Bio-Rad, Hercules, Calif.). Three size selections were performed to increase the average insert size of the BACs, and to eliminate small DNA fragments trapped in the HMW fraction. The first PFGE size selection was performed for 18 h at 11°C with a constant pulse time of 90 s, at a 120°X angle and 6 V/cm. The second size selection was performed for 12 h at 11°C with an initial pulse time of 1 s, a final pulse time of 10 s, at a 120° angle and 6 V/cm. The third size selection was performed for 8 h at 11°C with constant pulse time of 5 s, at a 120° angle and 4 V/cm.

For ligation, the molar ratio of the vector to the insert DNA was 4:1. The whole ligation reaction mix was diluted to a one-third compared to a normal BAC cloning reaction (Frijters et al. 1997). After a 24-h incubation at 11°C, ligated DNA was transformed by electroporation into the *E. coli* strain ElectroMAX DH10B using a Gibco BRL Cell Porator and Voltage Booster system (Gibco BRL, Grand Island, N.Y.). The settings were 360 V, 330 μF capacitance, low ohm impedance and fast charge rate; the voltage Booster was adjusted by setting the resistance to 4,000 ΩW. Recombinant transformants were selected on a LB (Luria-Bertani) agar plate containing 15 mg/l tetracycline, 0.5 mM IPTG and 50 μg/ml X-gal. After a 24-h incubation at 37°C, individual white colonies were isolated into 384-well plates containing LB freezing medium. After incubation at 37°C for 14 h, the plates were stored at –80°C (Zhang et al. 1996). The average insert size was assayed by a simple alkaline lysis DNA mini-preparation (Sambrook et al. 1989), restriction digestion with *Not*I to free the DNA insert from the vector and size separation by PFGE (Frijters et al. 1997).

Large-insert library screening

A Biomek 2000 Automated robot (Beckman Coulter, Fullerton, Calif.) was used to spot the clones of the large-insert library onto 12 × 8-cm Hybond N+ filters (Amersham-Pharmacia, Piscataway, N.J.) of so that a filter would contain two copies from each colony of four 384-well micro-titer plates. The entire library of 38,400 *Bam*HI and 38,400 *Hin*dIII clones was inoculated onto 50 filters. The inoculated filters were placed on lids of 96-well micro-titer plates containing LB agar and 25 μg/ml tetracycline and incubated at 37°C overnight. When the colonies reached a size of 2–3 mm in diameter, the filters were prepared for hybridization (Zhang et al. 1996) with random hexamer [32P]-labeled DNA probes (Sambrook et al. 1989).

Subcloning of the Bng122 clone

About 25 ng of a hybridizing Bng122 clone was digested with 5 U of *Sau*3A enzyme for 1 h at 37°C, the restricted DNA was ligated into *Bam*HI-restricted dephosphorylated pGEM3 vector for 3 h at 37°C. The ligated DNA was electroporated into *E. coli* DH10B strain ElectroMAX. Transformants were selected on media containing ampicillin, IPTG and X-gal. About 700 recombinant clones and 15 non-recombinants were individually isolated into three plastic Petri dishes (100 × 15 mm) containing LB agar with 12.5 μg/ml ampicillin. After overnight incubation at 37°C, colonies were lysed on the Nylon membrane (Nizetic et al. 1990).

Bng122 subclones screening

About 100 ng of each oligomeric repeat primer (AT)10, (ATT)10, (AAT)10, (CT)10, (AG)10, (CT)10, (AC)10, (CG)10, (CC)10,(GG)10, was labeled by phosphorylating the 5′ end with 3 μl γ−[32P] ATP (92.5 × 10 Bq 3000 Ci/mmol) and 10 U of T4 Kinase (Pharmacia, Piscataway, N.J.) for 30 min at 37°C. The labeling reaction was stopped by heating at 68°C for 15 min. The filters were prehybridized in a solution of (5 × SSC, 0.5% SDS, and 5 × Denhardt's) for 2 h at 39°C and then hybridized overnight at 39°C in a prehybridization solution that contained the labeled probe. After hybridization, the filters were washed twice for 20 min at 37°C (in 2 × SSC, 0.2%SDS) and then twice for 1 h at 37°C (in 1 × SSC, 0.1%SDS).

DNA sequence analysis

Sequence determinations were performed by the dideoxy chain-termination method using an ABI big dye cycle sequencing kit and an ABI377 automated DNA sequencer.

Pool and super-pool of the two libraries

Pools and super-pools of mixed DNA from 384 and 1,536 clones, respectively, were constructed. Each 384-well micro-titer plate in the library was replicated into a plate containing LB agar and 25 μg/ml tetracycline. After overnight incubation at 37°C, cells from each plate were suspended (pooled) with 5 ml LB into one tube. Plasmid DNAs were isolated by alkaline lysis (Sambrook et al. 1989) and were stored at –20°C in microfuge tubes. Pools were constituted from the DNA from one 384-well plate and superpools from the DNA from four 384-well plates.

Microsatellite markers screening of the two libraries

The microsatellites primers were labeled by phosphorylating the 5′ end with 5 μl γ-[32P] ATP (92.5 × 10 Bq 3000 Ci/mmol) for 30 min at 37°C with 10 U of T4 Kinase (Pharmacia, Piscataway, N.J.). Radioactive polymerase chain reactions (PCR) (Meksem et al.1999) were performed with pool and subpool clone DNAs and with genomic DNA from our mapping population ($F_{5:13}$ recombinant inbred lines that segregate for SCN and SDS resistance). The PCR products were separated by electrophoresis on a 5% (w/v) polyacrylamide denaturing gel.

Fingerprinting of the large-insert clones for physical mapping

Individual clone DNAs were fingerprinted using the fingerprinting kit developed for large-insert bacterial clones (the Fpase kit; H.-B. Zhang and Q. Quanzhou, Invention no. TAMUS1228). The DNAs were extracted in 96-well plates, mixed with the enzyme mix Fpase I (Fingerprinting enzyme mix I), the Fpase I buffer and [32P]-dATP and incubated at 37°C for 2 h. The reactions were stopped with the DNA sequencing gel loading dye [98% (w/v) deionized formamide, 0.3% (w/v) bromophenol blue, 0.3% (w/v) xylene cyanol and 10 m*M* EDTA, pH 8.0]. Fragments were denatured at 95°C for 5 min and subjected to electrophoresis on a 4% (w/v) denaturing polyacrylamide gel. The gel was dried onto 3MM blotting paper and used to expose X-ray film for 12 h (Quanzhou and Zhang 1998).

Results

Construction of the Forrest large-insert DNA libraries

Two soybean large-insert libraries were constructed from cv. Forrest: one library from microbead DNA partially digested with *Bam*HI and a second library from plug DNA partially digested with *Hin*dIII. The size of the digested fragments ranged from 100 kb to 600 kb. After the first PFGE, the fraction of gel containing the 150- to 350-kb DNA fragments was transferred to a new gel. Fragments over 350 kb were excluded because after ligation such clones are very difficult to transfer intact into bacterial cells. Two more rounds of size selection by PFGE were performed. The second PFGE was performed to remove the smaller

DNA molecules (50–90 kb) trapped within the large fragments after the first size selection. The third PFGE concentrated the DNA. About 150 µl of size-selected DNA fragments (at 4 ng/µl) was harvested after the third size selection. About 6,000 recombinant clones were obtained from each single electroporation representing 1,250–1,500 recombinant clones from each nanogram of ligated DNA.

Analysis of 100 randomly selected clones showed that the average insert size was 125 kb (fig. 1). The V41 vector contains four *Not*I restriction enzyme sites generating three vector fragments after *Not*I digestion. The two *Not*I sites flanking the multicloning site allow the release of the insert. The largest insert size was 260 kb and the smallest 60 kb (fig. 2). About 60% of the clones contained inserts between 100 and 150 kb (fig. 1). Clones with no insert were not found. Therefore, both libraries contain 38,400 inserts each, with an average size per insert of about 125 kb.

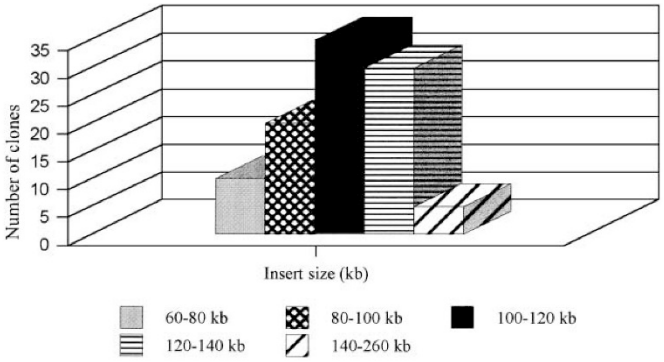

Figure 1. Insert size distribution of 100 clones randomly taken from the Forrest soybean *Bam*HI and *Hin*dIII large-insert DNA libraries. The insert sizes of over 50% of the clones are larger than 100 kb. The average insert size in the Forrest *Hin*dIII and *Bam*HI libraries is about 125 kb.

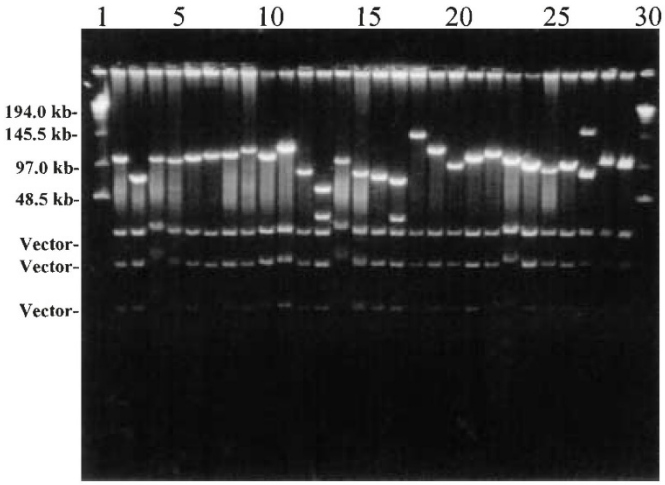

Figure 2. PFGE analysis of 28 randomly picked clones of the soybean Forrest *Hin*dIII and *Bam*HI large-insert DNA libraries. The common bands across all the lanes are the pCLD04541 vector fragments. *Lanes 2–29* Large-insert DNA clones, *lanes 1* and *30* λ-concatemers DNA size marker.

Representation of the large-insert library for the soybean genome

The soybean genome is estimated to contain 1.11 × 106 kb (Arumuganathan and Earle 1991). Therefore, together the two libraries from Forrest should represent almost 9.5-fold the haploid genome DNA content, and each library should represent about 4.7-fold the haploid genome DNA content. The probability of finding at least one copy of a unique sequence in each library should be around 99% (Clarke et al. 1976). To test this estimate, we screened each library with several different probes for unique sequences by colony hybridization and PCR amplification. The library was constructed from cv. Forrest, which contains genes for resistance to the soybean cyst nematode (Chang et al. 1997). Most of the probes used were microsatellite and sequence-characterized amplified region (SCAR) markers closely linked to *Rhg*1 on linkage group G and *Rhg*4 on linkage group A2, two loci that confer resistance to the soybean cyst nematode SCN (Meksem et al 1999, Matthews et al 1998). Between 2 and 8 BAC clones were detected with each probe (table 2). Therefore, the chromosomal regions carrying the two unlinked SCN resistance loci were significantly represented. The screening of both libraries with the two different sets of probes showed that the soybean genome is represented in both libraries. The screening of the two libraries also showed variation in the number of positive clones between the *Bam*HI and the *Hin*dIII libraries. For example, the *BamHI* library contained 5 positive clones to the BARC-Satt 309 probe, and the *Hin*dIII library contained only 2 positive clones for the same probe (table 2).

Table 2. Large insert DNA clones identified by probes closely linked to SCN (1, 2, 3, 4, 5, and 6), SDS (1, 4, and 5) and RSV1 (7) disease resistance genes and to the NTS (8) gene in soybean. NS: Non screened

		Number of clones identified			
Marker	Linkage group	*Bam*HI library	*Hin*dIII library	Total size (kb)	Mean insert
1 BARC-Satt 309	G	5	2	7	120
2 BARC-Satt 275	G	3	6	9	95
3 BARC-Satt 163	G	2	7	9	123
4 SIUC-Sat122	G	5	3	11	115
5 Bng122	G	5	2	7	110
6 pBLT65	A2	10	7	17	130
7 NBS5	F	6	NS	6	85
8 PA381-1	H	3	NS	3	80
Total		26	33	59	107.25

Ratio of chloroplast (cp) DNA clones in the large-insert library

Since the two large-insert libraries were constructed from leaf DNA, clones containing cpDNA were expected to be present in both. To estimate the percentage of clones containing cpDNA we hybridized ten filter membranes containing 15,360 clones from *Bam*HI and *Hin*dIII library with the soybean *psb*A cpDNA probe. The result indicated that 0.3% of the clones contain that chloroplast gene (data not shown).

Targeted microsatellite

Colony hybridization of the *Hin*dIII library with the Bng122 probe identified 6 clones. One clone (HRB122–5) with an insert size of about 110 kb was digested with *Sau*3 A and subcloned into the pGEM-3Z vector. About 700 recombinant clones were isolated, transferred onto Nylon membrane and hybridized with the oligomer repeats. Of nine hybridizing plasmids that were sequenced, six candidate microsatellites were identified on four plasmids. The candidate microsatellites detected were four to eight repeats of AT, ATT, CG, CTT or CT. Primers were designed to flank the four AT repeat within the hybridizing plasmid clone S122–5AT (fig. 3), and the fragment was shown to be polymorphic in the Essex × Forrest mapping population (fig. 4). The developed microsatellite marker was named SIUC-Sat122. Linkage analysis indicated that there were two recombinants among 200 informative meioses between the Bng122 marker and the SIUC-Sat122. These two markers were both found in the same 110-kb insert clone.

AATTTTTTATATAAGTTGCAAAATTTAGGGACTTATTTATTATTAAATTATTTG
TAGGGACTAATTTATCATATTTTTTGTATATTCAGGAATTAAATTTAATTTTTC
ATCCTTCAATACTAACTTATTAACGTTTCACATTTTCAAAGACGAGTCTAGCT
ATTTATAATTTTTTTTTCCTAAAATATATTTTTTGTCCTCATAAATATGAAAATA
TTTAAAATTCGTTCCTAATTTTTTTTTTCAAAGCATCTTTCCTT**CTCACAAAATT**
GAAATGTATCATTTTTTTTTGTTCAAAAGTTTAAATAAATTTGAACCTAATAT
GACATTTTATATCGGTTATACATATAACTGATATAAACATCAAGTTTTTTATA
TCAATGATACCTATAACTGATATCAAATGTGACAAT**TATATATATA**ATTAAT
GTAAAAAAGTCATAAATATAATTTATTTTGAGTCAAAAAATAATATATTTTAA
TT**ATTTTGAAGATGAAAAAGG**ATAAATTTAAAACGTTTGTGTGAGATGAAA
AACTAGATGTTTTTTTTCCTGGTTTAAATGCAAAACCAATGCTATTTTATTTAA
ATTTTACCTTTTTTTTATAATTACNCCACCAAAAAACCGTTTGGTGTTACAAAT
TTGA

Figure 3. Sequence of the S122–5AT plasmid clone. The position of the microsatellite repeat is indicated in *boldface* and underlined; the positions of the SIUC-Sat122F and SIUCSat122R primers used for polymorphism test are indicated in boldface.

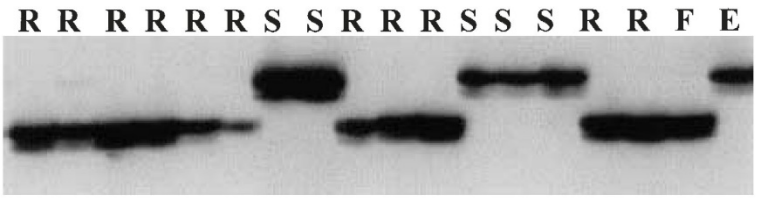

Figure 4. PCR amplification of genomic DNA from the resistant parent Forrest (*F*), the susceptible parent (*E*), and the F5:13 selected recombinant inbred lines segregating for SCN and SDS resistance with SIUCSat122 primers. The labeled PCR products were electrophoresed on a 5% polyacrylamide denaturing gel (*R* Resistant, *S* Susceptible) to SCN.

Fingerprinting of large-insert clones for soybean gene golfing

The fingerprinted clones produced an appropriate number of bands in the size range from 25 to 2,250 bp. About 25–50 labeled fragments were generated from a single soybean clone fingerprint; the bands visualized were distributed across the length of the lane (fig. 5), and every clone fingerprint was different from the next. When the soybean clones came from the same region of the genome, at least 4 bands were shared in common between two overlapping clones. Large-insert DNA clones recovered in chromosome landing and walking experiments were characterized by fingerprinting, and the fingerprints were used to build contigs close to *Rhg1* (fig. 5, lanes 1–8) and *Rhg4* (fig. 5, lanes 9–16).

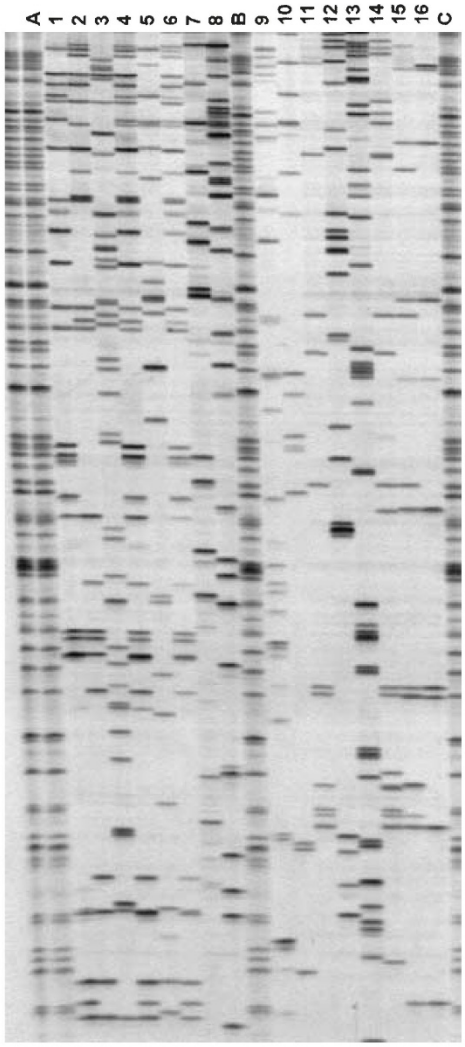

Figure 5. Fingerprint of the soybean large-insert DNA clones recovered in chromosome landing and walking experiments close to *Rhg1* (*lanes 1–8*) and *Rhg4* (*lanes 9–16*). *Lanes A, B*, and *C* Size standard marker (*λ-Sau*3 A)

Discussion

The binary V41 vector used to construct the soybean Forrest large-insert DNA library is a vector system that may accelerate the isolation of agronomically important genes (SCN, SDS, yield) that have been mapped in cv. Forrest (Chang et al. 1997; Njiti et al. 1997). V41 was initially designed for plant transformation via *Agrobacterium* (Jones et al. 1992) and has been successfully used to transfer DNA fragments of about 20 kb into plants (Bent et al. 1994; Brommonschenkel and Tanksley 1997).

Genome representation is a critical requirement for large-insert libraries during positional cloning and physical mapping. The combined *Bam*HI and *Hin*dIII libraries from Forrest contained about 8 positive clones per probe tested, close to the theoretical representation of 9.5 haploid soybean genomes. Representation was confirmed for four genomic regions during the screening of the library with the microsatellite markers BARC (-SATT309, -SATT275, -SATT163) and SIUC-SAT122, the RFLP probes Bng122, NBS5 and pA381, and the SCAR developed from the pBLT65 RFLP marker. Genome wide representation will be confirmed during physical map construction and genetic map integration of the soybean genome. Therefore, the libraries are useful resources for the soybean and legume research community (Men et al 1999).

The screening of the two libraries with the same probe rarely identified the same number of clones from both libraries. Similar variations were reported for lettuce (Frijters et al. 1997). Variations may be due to many factors, including sequence heterogeneity, the difference in GC-content of the recognition sites of the cloning enzymes (*Bam*HI, *Hin*dIII) and the nonrandom distribution and sampling of the DNA fragments during cloning. Large-insert libraries constructed with different restriction enzymes may be more suitable for assembling long continuous contigs from overlapping clones than a single library constructed with one enzyme due to heterogenous distribution of the sequences, repeats and GC-content in complex genomes.

Only 8% of the soybean large-insert clones contain inserts with internal *Not*I sites (fig. 2). In comparison, the rice (Zhang et al. 1996) and sorghum (Woo et al. 1994) BAC libraries contain more than three *Not*I sites every 150 kb. *Not*I has a GC-rich recognition sequence, whereas the soybean genome is AT-rich (Marek and Schoenmaker 1997).

Increasing the number of PFGE size selections has been suggested to decrease the large-insert cloning efficiency (Woo et al. 1994; Frijters et al. 1997; Vinatzer et al. 1998). To avoid DNA degradation, after the first size selection we did not melt down the fraction of gel containing the 150- to 350-kb restricted DNA when it was transferred to the second gel where the next two rounds of selection were performed. A partial Forrest soybean large-insert DNA library made with two size selections but with melting the agarose between gels contained only 100–150 recombinant clones per transformation.

An efficient purification and dephosphorylation (Danesh et al. 1998) of the vector before ligation was critical for a high transformation efficiency (5,000–6,000 recombinant clones/transformation). We found that the purity of the vector was very critical for high efficiency transformation and for reducing the number of clones with no insert (Marek et al. 1997; Kim et al. 1996). Three purification steps were performed to purify the vector in order to eliminate linearized plasmid and single-stranded and bacterial chromosomal DNAs that decrease the efficiency of large-insert DNA cloning and increase the number of empty vector clones.

The amount of cpDNA in the two soybean large-insert libraries (0.3%) was considerably less than rice, 0.94–2.89% (Zhang et al. 1996). Incubating the soybean plant for 72 h in dark and extending the number of nuclear pellet wash steps to six may have reduced cpDNA contamination.

The genetic distance between the Bng122 marker and the SIUC-Sat122 marker was 1 cM, corresponding to less than 110 kb. A ratio of physical to genetic distance of less than 110

kb/cM could be deduced for this interval, in agreement with Danesh et al. (1998). However, the ratio of physical to genetic distance will vary greatly across short distances and in separate progeny populations.

The two soybean libraries were constructed from Forrest, the parent of several mapping populations in our laboratory. These populations segregate for many traits of agronomic importance (Chang et al. 1997; Njiti at al. 1997; Meksem et al. 1999). Our recombinant inbred line (RIL) and near-isogenic line (NIL) mapping populations will provide an immortal and invaluable framework for soybean genome research. The libraries constructed from one parent of such a mapping populations will provide a common framework for physical mapping of the soybean genome. Finally the libraries can be used to clone genes by position and to assemble long-range overlapping clones for physical mapping of the soybean genome.

Large-insert DNA libraries fingerprinting for gene golfing

Fingerprinting approaches seek to produce fragments having sufficient overlaps so that a continuous path can be constructed by passing through a series of overlapping nearest neighbors. Discovering the original order of the cloned fragments is then dependent on detecting the necessary overlaps. The generation of data that permits the recognition of clone overlap, even though the clone fingerprints were not present on the same gel or generated in the same time period, is a key challenge in a fingerprinting methodology for gene golfing (Zhang and Wing 753 1997). The necessity for high throughput has made preparation of DNA in 96-well format a prerequisite. The small quantities of DNA achievable with low-copy number vectors in this format required highly sensitive detection methods.

Data generated by the fingerprint approach using Fpase (Quanzhou and Zhang, Invention no. TAMUS1228) were of higher quality than those used to construct the yeast, *C. elegans* and human physical maps (Coulson et al. 1986; Wong et al. 1997; Marra et al. 1997). Fpase methods combine the high throughput needed for DNA extraction, high sensitivity detection, adequate reproducibility and drastically reduced variation among migrating bands of identical sequences. The Fpase fingerprint method enabled contig assembly across centromeric repeat regions and can distinguish homoeologous sub-genomes in diploid rice (Chang et al. 1999). These features are essential for physical map construction in complex genomes. Advantages offered by this approach include data that are comparatively free of artifacts, compatibility with pre-existing software developed at the Sanger Center (Cambridge, UK) and the high throughput necessary for complex genome analysis. Integration with the microsatellite genetic map will allow the soybean genetics community to rapidly isolate genes underlying traits of agricultural importance (http://www.coalab.siu.edu/genome/).

The library and BAC contigs are available for collaborative purposes.

Acknowledgments – We thank Amanda Mulch, Nichole Long, Artem E. Men, Alec Hayes and Wang AL-Guo for technical assistance. The critical reading of the manuscript by Dr. M. Javed Iqbal and Boris Vinatzer is appreciated. This work was supported in part by the Illinois Soybean Board (ISPOB), the United Soybean Board (USB), and the National Science Foundation (NSF) Award Number 9872635.

References

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. Science 258: 1353–1355

Ashfield T, Danzer J, Held D, Clayton K, Keim P, Maroof S, Webb DM, Innes RW (1998) *Rpg1*, a soybean gene effective against races of bacterial blight, maps to a cluster of previously characterized disease resistance genes. Theor Appl Genet 96: 1013–1021

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–219

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science 265: 1856–1860

Bent E, Johnson S, Bancroft I (1998) BAC representation of two low-copy regions of the genome of *Arabidopsis thaliana*. Plant J 13: 849–855

Bonnema G, Hontelez J, Verkerk R, Zhang YQ, van Daelen R, van Kammen A, Zabel P (1996) An improved method of partially digesting plant megabase DNA suitable for YAC cloning: application to the construction of a 5.5 genome equivalent YAC library of tomato. Plant J 1996 9(1): 125–133

Brommonschenkel SH, Tanksley SD (1997) Map-based cloning of the tomato genomic region that spans the Sw-5 tospovirus resistance gene in tomato. Mol Gen Genet 256: 121–126

Coulson A, Sulston J, Brenner S, Karn J (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 83: 7821–7825

Chang SJC, Doubler TW, Kilo V, Suttner R, Klein J, Schmidt ME, Gibson PT, Lightfoot DA (1996) Two additional loci underlying durable field resistance to soybean Sudden Death Syndrome (SDS). Crop Sci 36: 1684–1688

Chang SJC, Doubler TW, Kilo V, Suttner R, Klein J, Schmidt ME, Gibson PT, Lightfoot DA (1997) Association of field resistance to soybean Sudden Death Syndrome (SDS) and cyst nematode (SCN). Crop Sci 37: 965–971

Chang YL, Tao Q, Wang J, Scheuring C, Meksem K, Zhang HB (1999) A large-scale plant transformation- and genome sequence-ready physical map of the *Arabidopsis thaliana* genome. Plant & Anim Genome VII. San Diego, W: 29 pp 37

Clark L, Carbon J (1976) A colony bank containing synthetic ColE1 hybrids representative of the entire E. coli genome. Cell 9: 91–99

Concibido VC, Lange DA, Denny RL, Orf J, Young ND (1997) Genome mapping of soybean cyst nematode resistance genes in Peking, PI90763, and PI88788 using DNA markers. Crop Sci 37: 258–264

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DJ, Chung J, Specht JE (1999) The integrated map of the soybean genome. Crop Sci 39: 1464–1490

Danesh D, Penuela S, Mudge J, Denny RL, Nordstorm H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96: 196–202

Diers BW, Keim P, Fehr W, Shoemaker RC (1992) RFLP Theor Appl Genet 83: 608–612

Frijters ACJ, Zhang Z, Van Damme M, Wang G-L, Ronald PC, Michelmore RW (1997) Construction of a bacterial artificial chromosome library containing large *Eco*RI and *Hin*dIII fragments of lettuce. Theor Appl Genet 94: 390–399

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. Science 269: 843–846

Hamilton C, Frarey A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci USA 93: 9975–9979

Hartwig EE, Epps JM (1973) Registration of Forrest soybeans. Crop Sci 13: 287

Jones JDG, Shlumukov L, Carland F, English J, Scofield SR, Bishop GJ, Harrison K (1992). Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Res 1: 285–297

Keim P, Diers BW, Olson T, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126: 735–742

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high density genetic map of soybean based upon AFLP markers. Crop Sci 37: 537–543

Kilo V, Abu-Thredeih J, Suttner R, Chang SJC, Gibson PT, Lightfoot DA (1997) Coinheritance of resistance to Sudden Death Syndrome (SDS) and Cyst Nematode (SCN) in soybean. Soybean Genet Newsl 24: 126–127

Kolchinsky AD, Landau-Ellis D, Gresshoff PM (1997) Map order and linkage distances of molecular markers close to the supernodulation (*nts-1*) locus of soybean. Mol Gen Genet 254: 29–36

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma TA (1993) Genetic map of soybean (Glycine max L,) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor Appl Genet 86: 901–906

Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996) Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet 92: 516–523

Liu YG, Shirano Y**,** Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning**.** Proc Natl Acad Sci USA 96: 6535–6540

Mansur LM, Lark KG, Kross H, Oliveira A (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L,). Theor Appl Genet 86: 907–913

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40: 420–427

Marra MA, Kukaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterson RH (1997) High throughput fingerprint analysis of large insert clones. Genome Res 7: 1072–1084

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262: 1432–1436

Matthews BF, MacDonald MH, Gebhardt JST, Devine TE (1998) Molecular markers residing close to the *Rhg4* locus conferring resistance to soybean cyst nematode race 3 on linkage group A of soybean. Theor Appl Genet 97: 1047–1052

Meksem K, Doubler TW, Chancharoenchai K., Njiti V, Chang SJC, Rao Arelli AP, Cregan PE, Gray LE, Gibson PT, Lightfoot DA (1999) Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines. Theor Appl Genet 99: 1131–1142

Men EA, Vasile CA, Anderssen I, Bhattacharyya MK, Meksem K, Lightfoot D, Gresshoff PM (1999) Progress in genomic analysis of the soybean supernodulation mutation *nts-1*. Plant & Anim Genome VII, San Diego, P: 79 pp 75

Mian MAR, Shipe ER, Alvernaz J, Mueller JD, Ashley DA, Boerma HR (1997) RFLP analysis of chlorimuron ethyl sensitivity in soybean. J Hered 88: 38–41

Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994) The *Arabidopsis thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78: 1089–1099

Njiti V, Schmidt C, Schmidt M, Lightfoot DA (1997) Mapping loci underlying yield in Illinois. Soybean Genet Newsl 24: 136–138

Nizetic D, Drmanac R, Lehrach H (1990) An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10, 11 bases) oligonucleotides to cosmids. Nucleic Acids Res 19: 182

Quanzhou Tao, Zhang Hong-Bin (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. Nucleic Acids Res 26: 4901–4909

Rong Di, Purcell V, Collins GB, Ghabrial SA (1996) Production of transgenic soybean lines expressing the bean pod mottle virus coat protein precursor gene. Plant Cell Rep 15: 746–750

Salimath SS, Bhattacharyya MK (1999) Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the *Rps1*-k disease resistance gene. Theor Appl Genet 98 5: 712–720

Sambrook J, Maniatis T, Fritsch EF (1989) Molecular cloning: a Laboratory manual, Cold Spring Harbor Laboratory Press, New York

Shoemaker RC, Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. Crop Sci 35: 436–446

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, and Boerma HR (1996) Genome duplication in soybean (*Glycine* Subgenus *Soja*). Genetics 144: 329–338

Singh RJ, Klein TM, Mauvais CJ, Knowlton S, Hymowitz T, Kostow CM (1998) Cytological characterization of transgenic soybean. 96: 319–324

Song WY, Wang GL, Chen L, Kim HS, Pi LY, Gardner J, Wang B, Holsten T, Zhai WX, Zhu LH, Fauquet C, Ronald PC (1995) A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science 270: 1804–1806

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JD (1995) Molecular genetics of plant disease resistance. Science 268: 661–667

Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR (1997) DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. Theor Appl Genet 95: 664–670

Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 2: 63–68

Vierling RA, Faghihi J, Ferris VR, Ferris JM (1996) Association of RFLP markers with loci conferring broad-based resistance to the soybean cyst nematode (Heterodera glycines). Theor Appl Genet 92: 83–86

Vinatzer BA, Zhang H-B, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. Theor Appl Genet 97: 1183–1190

Wang G-L, Holsten TE, Song W, Wang H-P, Ronald PC (1995) Construction of a rice bacterial artificial library and identification of clones linked to the *Xa21* disease resistance locus. Plant J 7: 525–533

Webb DM, Baltazar BM, Rao-Arelli AP, Schupp J, Clayton K, Keim P, Beavis WD (1995) Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437,654. Theor Appl Genet 91: 574–581

Wong GK, Yu J, Thayer EC, Olson MV (1997) Multiple-complete-digest restriction fragment mapping: generating sequence ready maps for large scale DNA sequencing. Proc Natl Acad Sci USA 94: 5225–5230

Woo S-S, Jiang J, Gill BS, Paterson AH, Wing R (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res 22: 4922–4931

Yang D, Parco A, Nandi S, Subudhi P, Zhu Y, Wang G, Huang N (1997) Construction of bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4-specific RFLP markers in rice. 95: 1147–1154

Yu YG, Buss GR, Saghai Maroof MAS (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 1996 93: 11751–11756

Zhang H-B, Wing RA (1997) Physical mapping of the rice genome with BACs. Plant Mol Biol 35:115–127

Zhang H-B, Zhao XP, Ding XD, Paterson AH, Wing RA (1995) Preparation of megabase-sized DNA from plant nuclei. Plant J 7: 175–184 Zhang H-B, Choi S, Woo S-S, Li Z, Wing RA (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2: 11–24